Tina Memo No. 2003-011
Internal Memo

Tutorial: A Critical Analysis of Voxel Based Morphometry (VBM).

N.A.Thacker

Last updated
13 / 5 / 2008

Imaging Science and Biomedical Engineering Division,
Medical School, University of Manchester,
Stopford Building, Oxford Road,
Manchester, M13 9PT.

A Critical Analysis of Voxel Based Morphometry (VBM).

N. A. Thacker.

This document was written on request for the RA's and students in our research group. Its purpose is to summarise concisely a recent debate in the literature regarding the use of Voxel based Morphology in scientific studies. The work was originally presented as a tutorial, and it is not intended for journal publication as I do not believe it contains anything which has not been said before by others. The document may however prove useful, as a way of explaining to potential collaborators the reasons someone might not wish to use VBM in the course of a serious scientific study [1]. A section is included which examines the potential consequences of these issues for a practical experiment. This thought experiment is then compared with the claims made in results of a data analysis carried out by the original developers.

Sources

As this document is intended to summarise an existing debate I will draw heavily on others words, adding only my own comments in order to try to explain the potential consequences of these statements. It is recommended that you have these documents at hand and refer to them in the course of reading this review, particularly for the details of the analysis technique and results.

The main sources for this document are;

- **A Voxel Based Morphometric Study of Ageing in 465 Normal Adult Brains. C. D. Good, I, S. Johnsrude, J. Ashburner, R. N. A. Henson, K. J. Friston and R. S. J. Frackowiak. NeuroImage, 14, 21-36 (2001)**

  Description of the basic method and application to normal variation. This includes some results from global change and a description of both the NEW and OLD VBM. This therefore makes this paper in particular a good starting point for an evaluation. I also know for a fact that some groups have modelled their data analysis strategy on this paper.

- **Registration Error and Functional Image Analysis. F. L. Bookstein. (Workshop on Biomedical Statistics, Leeds 2001)**

  Mainly a critique of the mapping process and statistical methods. Though it is admittedly difficult to follow and foccussing heavily on Booksteins own suggestion for a solution. But the fundamental criticisms of VBM are all there, the paper should not be discounted just because you can't understand/don't like the rest of it. Bookstein criticises the use of VBM for both FMRI and structural analysis. I'm only going to consider the latter here.

- **Understanding the Bookstein Controversy. W. Crum, (IRC report 2002)**

  Essential reading if you want a clear concise coverage of the main topics. I would have emphasised more the expectations of the user community as it is their expectations of the method, which in my experience seems to cause the most problems.

2001 is quite distant now, and even before this controversy Alan Jackson and I had already decided to avoid the use of the technique and to develop alternatives. Some of our fears regarding methodology of quantitation have since been borne out in our comparisons of basic measurements (see Appendix A). Though we have had to revisit the decision several times since when the issue was raised by collaborators. From what I can remember of our discussions they seemed to cover much the same topics as this public debate. Obviously, the methodology has changed slightly since then with the introduction of what we

[1] Even if everyone else is using it.

will be calling the NEW technique, so I saw this analysis as an opportunity to catch up. What I learned during previous discussions was that published comments seemed to carry much more weight than our own independent analysis and opinions. This tutorial was therefore deliberately constructed around quotes from published literature, even though all of what is said should be considered entirely amenable to careful reasoning, backed with perhaps a small amount of experience.

Some Background.

What can we generally achieve with Medical Image Analysis?

There are two main reasons for wishing to analyse medical image data; the pursuit of science and decision support.

In the case of science, we want physical or biological interpretations of data (measurements have real dimensional properties). It is generally accepted that experiments should be repeated by independent scientists in order to confirm any conclusions. Using multiple research approaches is generally useful as a way of excluding mis-interpretation. For medical data analysis, genuine understanding of a disease is attempted.

For decision support, we are largely performing a categorisation, where repeatability of measurement is important but not definition. Therefore using a standardised approach might be considered essential. We can define a 'good' technique a one which is useful for diagnosis.

What do people appear to want from structural analysis of brain images?

The Holy Grail would seem to be: The ability to identify significant structural changes between groups of brains, for the purposes of proving specific biological hypotheses.

Current Limits and Basic Knowledge.

There are several factors surrounding the limits of what can currently be achieve using standard methods. Specifically;

- Atlas approaches, such as Taliarach co-ordinates, have no better than 15mm resolution.
- Grey matter patterns and sulcal structures do not generally line up between normal individuals at the 1cm level.
- Total brain volume seems to scale all structural measurements.
- We know that there are age related effects which will appear significant in comparison to disease.
- There are large topological structural differences between individuals, which seem to make trivial comparison impossible.

What is Voxel Based Morphometry?

The essentials of the algorithm come down to the following;

- Take two groups of brains (or a distribution across a parameter such as age).
- Align each individual at the voxel level using non-rigid registration.
- OLD VMB. Simply warp the images.
- NEW VBM. Encode the deformation field for each subject as tissue density changes in the normal space.
- Blur everything [2]

[2] Luckily we can ignore the consequences of this stage for most of the remainder of this discussion. I'll come back to it briefly at the end.

- Analyse the group level differences between aligned voxels in these data sets using conventional Gaussian statistics.
- Produce image output as significance maps which identify the detected differences.

This approach is distributed as part of the SPM software and has been applied to both FMRI and structural brain analysis. The technique has become widely popular in some areas and yet at the same time also come under close scrutiny and been the subject of severe criticism. The question we wish to address here is; Why?

Criticisms

Area of Criticism 1: The basic idea.

Some people have problems with a statement like... **There is a meaningful one-to-one mapping which can be defined at the pixel level between any pair of brains using a low parameter non-linear model.** Where "meaningful" refers to the fact that something useful can be extracted from the estimates. In fact the first thing to say is that deformation fields are not estimates of volumetric change when there are topological changes present.

The original authors say;

Good: **"The non-linear transformations used in this study do not attempt to match every gyrus in every brain, rather the aim is to accommodate global brain shape differences,"**

and

Ashburner + Friston : **"VBM has been designed to be sensitive to these (small scale structural) differences, while discounting positional and other large scale volumetric changes in gross anatomy."**

Given the possibilities of sulcal, topological and structural changes, there is a continuum of possible change, but the algorithm must make an important decision at some point to exclude some changes and model others.

Bookstein: **"The claim that this algorithm has discounted global shape differences is wholly metaphorical"** to which

Crum: **"I think this (Bookstein) is right."** Yes I think he's right as well. The OLD vbm would detect no changes at all in structural comparisons if the non-rigid registration were accurate in all brain to brain mappings. This is probably the main reason we decided not to use it several years ago.

It has been pointed out that the information for alignment only occurs at edge data and in order determine flow fields at all points some assumptions need to be made which determines how to interpolate between the evidence. Quite rightly people want to know what the specific interpolation function is and how it was chosen. Friston : paraphrased - " the non-linear function is what SPM does" in response to.... what model is used for non-linear registration?

Bookstein : **"By the usual rules of peer-reviewed science, findings reported in this way are to be presumed independent of any omitted details -"** Implication - as the results would alter depending upon the assumed non-rigid alignment model then the results are un-scientific (Ouch!).

Area of Criticism 2: The methodology.

Can we assume that the the parameters of a non-rigid registration can be can be reliably and automatically determined? For this, fully automatic and robust non-rigid co-registration is required. This should be immediately recognisable as a potential problem to anyone who has ever attempted to fit a function to a data set which does not have quite the right shape. There would be large numbers of local minima. None of them will actually summarise the contents of the data.

Bookstein : **"Wherever there is partial registration the true value of a (vector deformation) is inaccessible."** The mapping measured, whatever it is just isn't meaningful let alone reliable.

Until now all issues raised require a high degree of experience in the area in order to make a judgement. All are points for concern and likely to have those who design algorithms for a living shuffling uneasily in their seats, but some may wish to ignore them. So to be on firm ground we will put these issues to one side and look for criticisms which can be agreed by both sides of the debate.

Area of Criticism 3: The statistics.

Can deformation maps be statistically analysed to reliably determine areas of structural change?

Good: **"The statistics used to identify structural differences assume that the residuals after fitting the model are normally distributed."** Unfortunately, some of Booksteins analysis concerning stability around edge features would seem to suggest that there might be such problems here, and if he doesn't explicitly suggest this then I would. This paper claims that the Gaussian distribution has been checked in previous work by Friston + Ashburner. But aren't the distributions under test specific to each experiment? In areas like particle physics, Monte-Carlo's are used on a case by case basis to show that each conclusion is consistent with the observed data. Such a check here would require a significant additional step in the methodology.

Bookstein: **"The pattern of statistically significant changes will be spatially varying and a FUNCTION OF MISREGISTRATION as much as true geometric variation."**

Ashburner: admits VBM is - **"sensitive to systematic shape differences attributable to misregistration from the spatial normalisation step." (problem 1)**

Thus areas of significance can occur due to methodological error. Good et. al. also says there area problems due to segmentation. even the authors admit this is an issue so we will accept it as a valid criticism that we must take into account when interpreting results.

Area of Criticism 4: The science.

Can we assume that the characteristics of the mapping field, when encoded as density changes in a normalised tissue map, will be related to (some how) genuine tissue loss? these are the comments generated within the scientific debate.

Ashburner: **"Statistical tests do not protect against false negative results" (real differences are not necessarily detected) (problem 2)** A real and obvious limitation of the technique which everyone agrees on, on both sides of the debate.

Bookstein: **"The pattern of statistically significant changes will be SPATIALLY VARYING and a function of misregistration as much as true geometric variation."**

Good: There are - **"implications for the sensitivity of any morphometric technique to detect changes in regions of high variance." (problem 3)** Again, a clear and agreed statement of the limitation of morphometry. We MUST take this point seriously.

A Thought experiment.

Let us imagine that we intend to perform a scientific study to investigate brain volume changes between groups of individuals. There are a very limited number of possible outcomes from this study. Are any of them informed by the use of VBM?

We have a set of identified "statistically" significant changes, can we come to any conclusions regarding this set and any previous result? when:

- We saw a change but no change was expected. Can we reliably conclude that we have identified a new area of localised brain change? — No due to **problem 1**.
- We saw no change and no change was expected or we saw no change and change was expected. Can we exclude the possibility of change in a specific region? — No due to **problem 2**.

- We saw a change and a change was expected. Can we verify previous scientific observations regarding the areas of majority tissue loss? — No due to **problem 3**.

The observant will have noticed that we have just excluded all possible outcomes of the experiment. This is looking very bad for VBM, but we can still be proved wrong if we can find papers describing a scientifically useful conclusion in a practical use of VBM. So what results have the developers reported?

Analysis of Conclusions in Good et al.

Plots from Good et al. show globals changes, specifically the increase in measurement stability afforded by ICV normalisation. Age range covering 20-80 years. There were 450+ individuals, a very large cohort by any standards.

Regional results are summarised with carefully chosen phrases.

"... local areas of *relative* accelerated loss ... "

But can we define relative — (**problem 3**). Relative cannot mean relative tissue proportions, because we have already established that the method cannot do that. Relative here seems to be used in the context of "not absolute" and not "more than", and has presumably been included to get the paper past picky reviewers. It is understandable how people who are not experts in the algorithmic area may be miss-led by this statement into thinking that it is referring to relative tissue volume.

" ... areas exhibiting little or no age *effect* ... "

Define effect — (**problem 2**). The method certainly is incapable of putting any bound on volumetric change, so effect here just refers to the calculated statistic, nothing more. There may be tissue volume differences which do not reach significance due to the high topological variability of the surrounding structure.

Regional maps from Good et al. may seem to be suggesting areas of largest volume change, but look more closely at the Tables in Good et. al..

There are only Z scores in the summary tables, where are the (dimensional) quantifications of tissue change? It is unfortunate that P values and Z scores have become the lingua-franca of biological analysis, to the point that they can be quoted and accepted without any objective definition of what they are P values or Z scores of. A statement such as, the probability that a volumetric change has occurred in this area at a level greater than 10 % in the local 1cm cube region, would be considered rigorously scientific. Of course the authors, quite rightly, cannot report such numbers because the method does not deliver them. As we have already seen, no morphological approach could do this. We must instead use an interpretation which people like Bookstein finds so un-informative and unscientific, something like "the probability of the null hypothesis of statistical similarity, assuming Gaussian distributions, between voxel groups following the specific non-rigid alignment as embodied in SPM". And as we have already discussed earlier "relative" means "not absolute" though perhaps "non-ordinal" would have been better.

Why are there two sets of results which don't agree (moderated and unmoderated) ("Old" and "New" VBM by Crum), and why is there no discussion of these differences? Significantly only the results from the NEW method are reported in the abstract and conclusions. Could it be that criticisms regarding partial registration had struck home and the authors already think that OLD results are meaningless? Either way it still seems rather strange to have two contradictory sets of results presented in the same paper without adequate explanation.

Was the modulated (NEW) result also corrected for ICV? I don't think it was, based on their discussion section, this is strange as normalisation for inter-cranial volume would be very sensible based upon generally accepted results in the literature, and should have improved the statistical stability/power of their methods.

When attempting to compare results to previous literature: There is a large preamble covering at length the contradictory nature of the existing literature and finally.....

Regional effects of age: **"The statistical model used in this paper did however not model the global amount of grey matter and the age range is substantially different, so it cannot be directly compared to our method."** This may have actually been referring more to the results from global grey matter change, but its in the regional section, and its the only relational statement I can find.

The age range in this (Good) paper is 20-80 years, what age range was covered in the other literature which makes it so difficult to do a comparison? 80-120 perhaps?

Sex differences : **"We find it difficult to compare the results of these studies with our own, particularly since the majority of these studies are confounded to a greater or lesser degree by methodological factors such as normalising for brain size, region of interest measurements, quality of structural images, and the use of manual segmentation techniques.** So 30 years of manual measurement of structures has really generated nothing which can be compared to VBM? Do manual segmentation techniques always give results which are unreliable? Isn't the real reason that all other work was attempting to quantify absolute tissue volume and VBM does not. And what about normalisation to ICV? Does this imply that new VBM does not do it? Perhaps a non-linear co-registration makes a definition of ICV un-definable some way.

Lets make this clear, I am not suggesting that the Good paper in any way makes claims which cannot be substantiated by the data. I am however saying that the language used in places may mis-lead the unwary into thinking that something has been achieved which has not. Specifically, quantitative estimation of tissue loss. If interpreted carefully all reported results are consistent with the **thought experiment** and I would say of extremely limited scientific value. I find the global tissue plots the most informative part of the paper, but unfortunately these are inconsistent with other published literature (see appendix A).

The Future.

Things continue to change and it may be reasonable to think that the above analysis is in on shifting sands. Perhaps SPM 2005 will fix all of this? [3]

Good et al. **"We are currently attempting to ... provide better grey/white matter contrast and segmentation."** (see Appendix A)

and : **" As high dimensional warping techniques advance to the stage where they can provide accurate mapping of small gyri in large subject groups in a time efficient and practical way, then some of these issues will be resolved."** Well the OLD method of VBM won't identify any changes at all if you do this.

Alan Evan's group is looking at trying to fix the statistics.

All of these only have implications for aspects of **methodology** and **problem 1** (statistics). Nothing can be done regarding **problem 2** (non-detection) and **problem 3** (non-ordinal significances).

Perhaps these should be regarded as features rather than problems. We can thereby entirely eliminate these problems by the age old technique of re-definition. Repeatedly in the Bookstein/Ashburner and Friston debate, Bookstein is accused of not understanding what is intended by the designers of VBM. I believe that this is because Bookstein believes that in order to answer scientific questions, you really need volumetric estimates. He has interpreted the intentended use of VBM accordingly and found it lacking. It's O.K. for the originators of the technique to correct him on this, but they must bear some of the responsibility for this mis-understanding. A misunderstanding which still persists in some areas, particularly amongst potential user groups.

If you don't like the **basic idea** then only conclusive experimental tests will change your mind. So, should we interpret agreements with expectation as validation of the method? The method, though flawed, is still analysing images of brains and even if it were completely random it would be be expected to throw up some results which researchers expect to see. For non-scientists, accumulation of such events might be seen as proof of validity. However, as the technique cannot be expected to identify all known differences this is a bit of a one sided confirmation. Unfortunately the work just has not been carried out to show that this technique will identify what it is supposed to, and expected results are not a substitute for having a well designed methodology. As I have mentioned previously, I would expect it to be necessary to back up any claims of anatomical variation with a Monte-Carlo study showing that the claimed change would result in the observed measurement.

[3] Interestingly, I have now (2007) started to see papers in the literature which cite this very tutorial and then claim that these issues have been addressed. A recent paper in BJR claims that this tutorial focusses on the accuracy of registration, and that this has been fixed. Not only do I consider this claim to be over-optimistic, I think they must have stoppped reading this document at page 2.

On a similar note, Bill Crum has suggested running some simple evaluation/repeatability studies within the IRC. For example, take a set of normal brains, modify their structure in specific anatomical locations, then compare this set with another group of normals and see if the original set of structural changes are identified. This would be a lot of work if it were to replicate the kinds of numbers used in Good et al. In addition, as the problems are expected to be location specific, we could prove that VBM does not work this way, but could not prove that it would be expected to work in all cases. Good et al. claim that tests have already shown the method to work, so the IRC should look again at these experiments before embarking on any of its own. Perhaps this is another case of careful wording and all of the evidence is already in the literature.

Not Another Problem!

Someone (not me) should also look at sensitivity issues, the method may be fundamentally incapable of identifying change in some areas.

Bookstein says that results anywhere near an edge should be discounted due to statistical instability and blurring extends the range of this problem. I say that the only place that alignment can be determined is at edges so if this is true it poses a real problem.

I started this analysis looking to see if my own perception of the problems had been reflected anywhere in the debate. In fact whenever I thought I had something new I'd eventually realise that it was burried somewhere in the existing debate. By the end spotting a problem that I could claim as original became a challenge. I have not seen the spatial accuracy of VBM discussed anywhere (Bill might know better), so here are my thoughts.

A change in the sensitivity of the significance measure of a factor of two over a distance of 2 SD of a Gaussian distribution of width 14mm (the smallest spatial scale measurable following the blurring process) will produce a systematic shift of 5mm. Larger sensitivity changes will produce proportionately larger effects. It is well known in the image processing literature that a narrow highly curved edge structure (eg: a corner) can move by distances of the order of the smoothing kernal, ie: 14 mm. Worst case scenario, significance changes reported at maxima may be 2cm away from the edges that generated the detected signal and because of the nature of brain structure these edges may easily be as far away again from the actual site of tissue loss. Despite the use of non-rigid coregistration in the analysis and presentation of results in high resolution images, it would be difficult to believe that reported localisations from the method could significantly surpass the accuracy of even an affine aligned Taliarach system. The method may also shift the apparent position of change from the original site of tissue loss due to interactions between edge structures, the blurring process and the spatially varying sensitivity. (possibly **problem 4**)

Conclusions about VBM.

Returning now to the initial discussion on the reasons for doing medical image analysis.

Can we use VBM for decision Support?

It just doesn't seem to have enough statistical power to separate individuals as normal topological changes are already far too large to allow any sensitivity to disease. Interestingly the requirement for large statistical samples would also exacerbate the problems of applying Gaussian statistical models to non-Gaussian data.

A low parameter deformable co-registration should not be expected to reliably find a biologically meaningful deformation between two brains.

Can we use it for Science?

VBM may confound the processes of mis-registration and true anatomical change to the extent that statistical conclusions cannot be trusted. VBM must be considered incapable of ordering the **volumetric** significance of identified changes. Changes in the assumed non-linear mapping function would change the quantity and significance of identified regions.

Bookstein: **"...this degree of ambiguity should be judged intolerable in any applied scientific context."** Ouch again!

The authors insistence that we should use SPM as the definition of the process contradicts it's use for scientific purposes (take a look at Appendix A if you still believe otherwise, scientific progress requires independent confirmation in order to eliminate systematic effects). This is however consistent with the general attitude of the authors and provides a highly convenient excuse for those who wish to analyse data without making an attempt to understand what they are doing.

It is expected by many to provide an automated replacement for manual measurement of structures. Clearly it isn't and we must all make efforts to try to explain this to potential users.

Worryingly, (and as a direct consequence) VBM seems to generate results which cannot be directly compared to any previous research results even by the people who developed it. All of these concerns MUST be enough to convince any who want to do rigorous science that this technique should be avoided. The quality of scientific results (if any) obtainable with the method would seem to be very poor in comparison to the amount of work involved in running these experiments. There must be better ways of making use of the same data.

Recently, discussions with those who use this technique has demonstrated they believe that despite its faults, VBM can be used to look for 'interesting' behaviour between groups of subjects. This is fine provided they remember that the most significant regions of change identified using VBM can not be interpreted as the most significant (volumetric) structural change. They must also be wary of the statistical validity (and long term standing) of their results. If experiments are ever conducted which undermine the statistical validity of **any** studies using this methodology, this will cast doubt on all publications using these techniques, regardless of how careful researchers personally feel they were with their own study. In light of the technical discussion here, one such experiment to look out for would be if there were to be multiple independant analyses of a common dataset, or population group.

Boookstein: **"The only way I know to represent this structure is ..., by explicitly modelling covariation of deformation parameters along with the principle components of image variation they induce."**

There are indeed other approaches to brain volume analysis which do not suffer the same problems of morphological approaches. In particular, techniques which measure dimensioned quantities directly, such as cortical thickness, might be expected to show more promise as scientific tools. This is an area thay we will investigate in future.

Appendix A

The work in this appendix was conducted much later than the original tutorial and will appear in more detail in a subsequent Tina memo (Paul Bromiley et.al.). It has been added to this document because it casts more light on the data presented in the Good et. al. paper.

The figure below shows a summary of published figures for the total cranial fluid volume as a fraction of inter-cranial volume [4]. Despite significant differences between definitions of the measured anatomical volume and measurement techniques there is remarkable agreement between most published results. The curve shown is a fit to our own data for 75 subjects between the ages of 20 and 80. The major discrepancies are for the published data in Good et. al (filled stars) which, given the size of study (460 subjects), should have provided the most statistically accurate estimates. In fact it is in disagreement with the expected anatomical values by 200 %. In order for this to happen this discrepancy must be compensated for by the associated parenchymal volume estimates (for example grey matter and white matter fractions both being under estimated by 9 %). The other anomalous result (open star) was also obtained using SPM, and given the quoted errors, is inconsistent with either Good et. al. or. the other papaers. These results would seem to imply serious methodological flaws in these papers.

We can only speculate as to the cause of these results, but we do know from our own experience that techniques which do not model partial volume processes appropriately can give very unstable segmentation estimates. Fitted parameters are adjusted to account for data which is not correctly described by the

[4] The potential value of cranial fluid measurements (as opposed to grey/while matter volumes), as a way of assessing the excess of atrophy above that expected by normal ageing, is often overlooked.

assumed pure tissue distributions according to the relative fraction of pure to partial volume data seen in each brain. This can produce mis-leading segmentations in fully automatic approaches without careful quality control. It was exactly these problems which led us to develop our own automatic techniques, based upon partial volume analysis (described in several documents on our web pages).

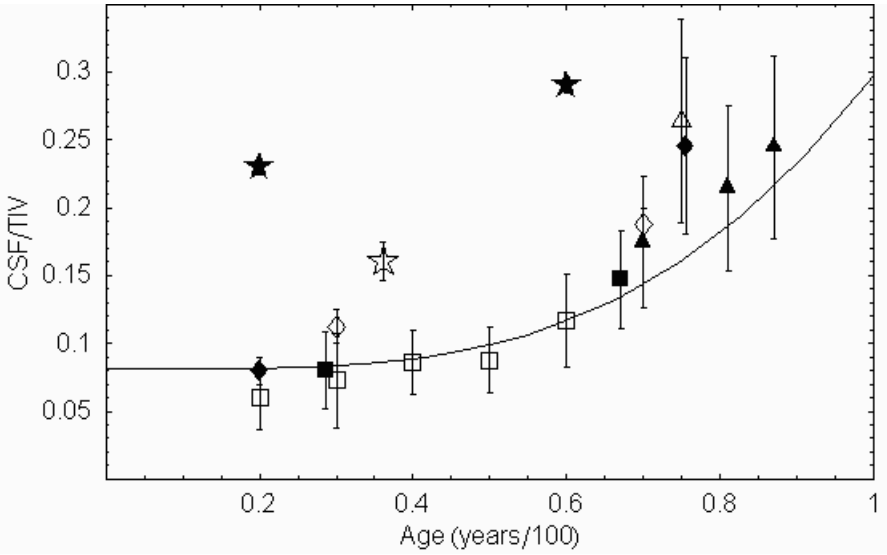

Figure 1: *Measurements of total CSF volume.*

Key

- (curve) TINA, partial volume analysis.
- (filled square) Gur, R.C., Mozley, P.D., Resnick, S.M., Gottlieb, G.L., Kohn, M., Zimmerman, R., Herman, G., Atlas, S., Grossman, R., Beretta, D., Erwin, R., and Gur, R.E., Gender differences in age effect on brain atrophy measured by magnetic resonace imaging. Proc. Natl. Acad. Sci. USA, Vol. 88 pp. 2845-2849, 1991.
- (open sqare) Blatter, D.D., Bigler, E.D., Gale, S.D., Johnson, S.C., Anderson, C.V., Burnett, B.M., Parker, N., Kurth, S., and Horn, S.D., Quantitative Volumetric Analysis of Brain MR: Normative Databse Spanning 5 Decades of Life. AJNR Vol. 16 pp. 241-251, 1995.
- (filled triangle) Mueller, E.A., Moore, M.M., Kerr, D.C.R., Sexton, G., Camicioli, R.M., Howieson, D.B., Quinn, J.F., and Kaye, J.A., Brain volume preserved in healthy elderly through the elevnth decade, Neurology Vol. 51 pp. 1555-1562, 1998.
- (open triangle) Coffey, C.E., Saxton, J.A., Ratcliff, G., Bryan, R.N., and Lucke, J.F., Relation of education to brain size in normal aging: Implications for the reverse hypothesis Neurology Vol. 53 pp. 189-196, 1999.
- (open diamond) Whitwell. J.L., Crum, W.R., Watt, H.C., and Fox, N.C., Normalisation of Cerebral Volumes by Use of Intracranial Volume: Implications for Longitudinal Quantitative MR Imaging AJNR Vol. 22 pp. 148-1489, 2001.
- (filled diamond) Courchesne, E.R., Chisum, H.J., Townsend, J., Cowles, A., Covington, J, Egaas, B., Harwood, M., Stuart, H., Press, G.A., Normal Brain Development and Aging: Quantitative Analysis at in Vivo MR Imaging in Healthy Volunteers. Radiology Vol. 216 pp. 672-682, 2000.
- (filled star) Good, C.D., Johnsrude, I.S., Ashburner, J., Henson, R.N.A., Friston, K.J., and Frackowiak, R.S.J., A Voxel-Based Morphometric Study of Ageing in 465 Normal Adult Human Brains NeuroImage Vol. 14 pp. 21-36, 2001.
- (open star) Chard. D.T., Parker, G.J.M., Griffin, C.M.B., Thompson, A.J., and Miller, D.H., The Reproducibility and Sensitivity of Brain Tissue Volume Measurements Derived from an SPM-Based Segmentation Methodology JMRI Vol. 15 pp. 259-267, 2002.